

구글 클라우드 넥스트 23의 주요 발표 내용



CONTENTS

디지털서비스 이슈리포트

- | | | |
|----|--------------------------------------|----|
| 01 | 구글 클라우드 넥스트 23의 주요 발표 내용 | 3 |
| | 한상기 테크프론티어 대표 | |
| 02 | 클라우드와 오픈소스를 활용한 로스앤젤레스 카운티 투표 시스템 사례 | 10 |
| | 윤대균 아주대학교 교수 | |
| 03 | 맥킨지 리포트: 2023년 중요 기술 동향과 전망 | 18 |
| | Senior Program Manager 김영욱 | |
| 04 | AI 기반 자율 운영 플랫폼 제공하는 비즈데이터 | 29 |
| | 도안구 테크수다 편집장 | |

본 저작물은 디지털서비스 이용지원시스템이 저작권을 보유하고 있습니다.

디지털서비스 이용지원시스템의 승인 없이 이슈리포트의 내용 일부 또는 전부를 다른 목적으로 이용할 수 없습니다.

01 구글 클라우드 넥스트 23의 주요 발표 내용

| 한상기 테크프론티어 대표

구글 클라우드는 아마존 AWS와 마이크로소프트 애저에 비해 3위의 기업이지만 인공지능 분야에서 발표하는 기술의 내용은 많은 기업에게 좋은 참고가 된다. 특히 클라우드에서 지원하는 인공지능 레이어의 기술 수준과 제공하는 서비스 범위에 제한이 있는 국내 클라우드 서비스 제공기업으로서는 매년 열리는 구글 클라우드 넥스트의 주요 발표 내용을 깊이 있게 살펴볼 필요가 있다.

이번 글에서는 발표 내용 중에 인공지능과 관련한 기술과 서비스에 초점을 맞춰서 정리해 보고자 한다. 특히 올해 발표 내용에는 인공지능과 관련하지 않은 내용을 찾기 힘들 정도로 대부분의 발표가 인공지능 관련 기술과 제품 및 서비스이며 이는 구글 클라우드가 어느 회사보다도 인공지능 영역에서는 앞서가고자 하는 의지를 보인다고 평가할 수 있다.

행사는 2023년 8월 29일부터 31일까지 샌프란시스코에서 열렸고, 내년 2024년 행사는 4월 9일부터 11일까지 라스베이거스에서 열릴 예정이다. 구글 클라우드의 마케팅 담당 부사장인 앨리스 웨이곤펠드가 올린 블로그에 따르면 주요 발표 항목은 161가지에 달했다고 한다.¹⁾



그림 1 구글 클라우드 넥스트 세션 카탈로그 이미지 [출처: 구글]

1) Google Cloud, "All 161 things we announced at Google Cloud Next '23 – a recap," Sep 2, 2023

디지털서비스 이슈리포트

선도적 인공지능 인프라 제공: TPU v5e와 A3 VM

구글은 이번에 선도적인 인공지능 인프라 기술인 TPU와 GPU와 관련한 제품 포트폴리오를 크게 개선한 중요한 사항을 발표했다. 먼저, 현재 프리뷰 버전으로 공개한 클라우드 TPU v5e는 비용 효율적이고 다재다능하며 확장성이 뛰어나다는 점을 강조했다. TPU v5e는 구글 쿠버네티스 엔진(GKE), 버텍스 AI, 파이토치, JAX, 텐서플로우와 같은 주요 프레임워크와의 통합을 제공하므로 사용하기 쉽고 익숙한 인터페이스로 시작할 수 있다. 또한 대규모 모델을 구동할 수 있는 엔비디아 H100 GPU 기반의 A3 슈퍼컴퓨터 VM을 다음 달에 정식 출시될 예정이라고 발표했다.

중대형 학습 및 추론에 필요한 비용 효율성과 성능을 제공하기 위해 특별히 설계한 TPU v5e는 TPU v4에 비해 LLM(거대언어모델) 및 생성형 AI 모델에 대해 비용당 최대 2배 높은 훈련 성능과 최대 2.5배 높은 추론 성능을 제공한다. TPU v4의 절반도 안 되는 비용으로 더 많은 조직에서 더 크고 복잡한 인공지능 모델을 훈련하고 배포할 수 있게 한다는 것이다.

비용 이점을 얻기 위해 성능이나 유연성을 희생할 필요 없이, 성능, 유연성, 효율성의 균형을 맞춘 TPU v5e 포드를 통해 최대 256개의 칩을 400Tb/s 이상의 총 대역폭과 100페타옵스의 INT8 성능으로 상호 연결할 수 있다. 또한 TPU v5e는 단일 슬라이스 내에서 하나의 칩에서 250개 이상의 칩에 이르는 8가지 가상 머신(VM) 구성을 지원해 다양한 LLM 및 생성형 AI 모델 규모에 적합한 구성을 선택할 수 있다.

스케일아웃 인프라를 위해서는 구글 쿠버네티스 엔진인 GKE에서 클라우드 TPU를 정식으로 제공함으로써 TPU를 더욱 간편하게 운영할 수 있게 되었고 GKE를 활용하여 클라우드 TPU v4뿐만 아니라 클라우드 TPU v5e에서 대규모 인공지능 워크로드 오케스트레이션을 관리함으로써 인공지능 개발 생산성을 향상할 수 있다고 주장한다.

또한 관리형 서비스의 단순성을 선호하는 조직을 위해서는 클라우드 TPU VM을 사용하여 다양한 프레임워크와 라이브러리로 훈련을 지원하는 버텍스 AI를 사용할 수 있다. 마지막으로, 학습 작업을 더 쉽게 확장할 수 있도록 멀티슬라이스(Multislice) 기술을 프리뷰로 공개해 사용자가 물리적 TPU 포드의 경계를 넘어 최대 수만 개의 클라우드 TPU v5e 또는 TPU v4 칩까지 인공지능 모델을 쉽게 확장할 수 있도록 지원한다.

멀티슬라이스를 통해 개발자는 단일 포드 내에서 칩 간 상호 연결(ICI)을 통해 또는 데이터 센터 네트워크(DCN)를 통해 여러 포드에 걸쳐 워크로드를 최대 수만 개의 칩까지 확장할 수 있다.

디지털서비스 이슈리포트

멀티슬라이스 기술은 최첨단 PaLM 모델 개발의 원동력이 되었고 이제 구글 클라우드 고객에게도 제공한다는 의미이다.

지난 5월에 발표한 엔비디아의 H100 텐서 코어 GPU를 기반으로 하는 A3 VM은 차세대 AI 워크로드 및 LLM을 학습시키고 서비스하기 위해 특별히 설계한 것으로 이전 세대에 비해 3배 빠른 학습과 10배 더 큰 네트워킹 대역폭을 통해 슈퍼컴퓨팅 기능에서 큰 도약을 이루었다.²⁾

A3 VM은 듀얼 차세대 4세대 인텔 제온 스케일러블 프로세서, VM당 8개의 엔비디아 H100 GPU, 2TB의 호스트 메모리를 갖추고 있다. 최신 엔비디아 HGX H100 플랫폼을 기반으로 구축된 A3 VM은 4세대 엔비디아 NVLink 기술을 통해 8개의 GPU 간에 3.6TB/s의 양방향 대역폭을 제공한다. A3 네트워킹 대역폭의 개선은 티타늄 네트워크 어댑터와 엔비디아 집단 통신 라이브러리(NCCL) 최적화를 통해 이루어졌으며 전체적으로 A3는 최첨단 인공지능 모델을 구축하고자 하는 혁신가와 기업에게 큰 도움이 될 것이라고 말한다. 이번에 발표한 하드웨어 인프라의 개선은 미드저니, 캐릭터.ai, 앤트로픽, 어셈블리AI 등에서 이미 활용하면서 그 성능에 대한 만족감을 표현하고 있다고 발표에서 언급하고 있다.

버텍스 AI 업데이트

버텍스 AI에서 생성형 AI를 지원한다는 발표는 6개월 전에 했다. 이를 통해 지난 4분기에 사용자가 15배 이상 늘었다는 것이 구글의 얘기이다.³⁾ 이번 행사에서는 버텍스 AI의 기능을 확장하고 기본 모델을 쉽게 실험 및 구축하고, 엔터프라이즈 데이터로 맞춤화하며, 개인정보 보호, 안전 기능 및 책임감 있는 AI가 내장된 애플리케이션에 원활하게 통합 및 배포할 수 있도록 역량을 강화했다는 발표가 있었다.

- 모델 가든에 추가한 새로운 모델: 새로 추가된 모델로는 메타의 라마(Llama 2) 및 코드 라마(Code Llama)와 기술 혁신 연구소(TII)의 팔콘(Falcon) LLM이 있으며, 앤트로픽의 클로드 2를 사전 발표했다. 이러한 발표를 통해 자사, 오픈 소스 및 타사 모델을 아우르는 엄선된 모델 컬렉션을 갖추고자 한다. 이제 모델 가든에는 100개 이상의 모델이 있다고 한다. 특히 어댑터 튜닝과 인간 피드백을 통한 강화 학습(RLHF)을 모두 지원하는 유일한 클라우드 제공업체라고 주장한다. 버텍스 AI에서 모델 가든의 역할은 아래 그림 2를 통해 확인할 수 있다.

2) Google Cloud, "Announcing A3 supercomputers with NVIDIA H100 GPUs, purpose-built for AI," May 11, 2023

3) Google Cloud, "Vertex AI extends enterprise-ready generative AI development with new models, tooling," Aug 30, 2023

디지털서비스 이슈리포트

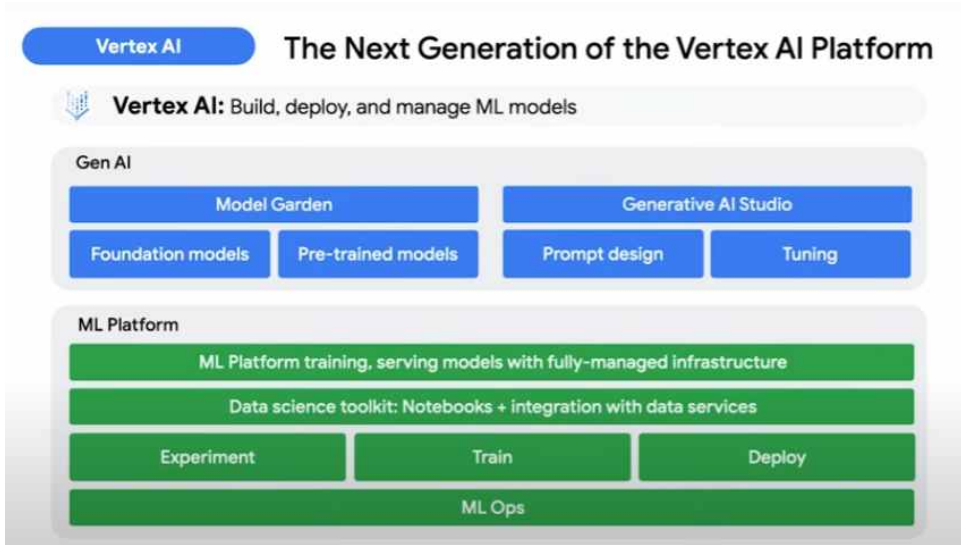


그림 2 버텍스 AI와 모델 가든

- 구글 자체 파운데이션 모델에 대한 업데이트를 통해 구글 딥마인드의 전문성을 제공하고, 더 높은 품질의 출력, 훨씬 더 큰 문서를 간단하게 분석할 수 있는 32,000개의 토큰 컨텍스트 창, 엔터프라이즈 데이터에 대한 그라운드링 기능 등을 갖춘 PaLM 2로 업그레이드된 것을 포함한다. 코드 생성 및 채팅용 모델인 코디(Codey)는 이제 25% 더 나은 성능을 제공하며, 이미지 생성용 모델인 이미젠(Imagen)도 향상된 이미지 품질을 제공한다.
- 구글 딥마인드의 신스아이디(SynthID)⁴⁾로 구동되는 이미젠의 새로운 디지털 워터마킹 기능
- 기업이 모델에서 더 많은 가치를 창출할 수 있도록 지원하는 새로운 도구. 여기에는 모델이 실시간 데이터를 검색하고 실제 작업을 수행할 수 있도록 지원하는 버텍스 AI 익스텐션과 다양한 소스에서 데이터 수집 및 읽기 전용 액세스를 제공하는 버텍스 AI 데이터 커넥터를 포함한다.
 - 익스텐션 기능을 사용하면 개발자는 인기 있는 엔터프라이즈 API에 대해 미리 빌드된 익스텐션 기능을 사용하거나 비공개 및 공용 API에 대한 자체 익스텐션 기능을 구축할 수 있다. 개발자는 이를 사용해 디지털 어시스턴트, 검색 엔진, 자동화된 워크플로와 같은 강력한 차세대 AI 애플리케이션을 구축할 수 있다.
 - 버텍스 AI는 빅쿼리, 엘로이DB와 같은 클라우드 서비스와 데이터스택스, 몽고DB, 레디스 같은 데이터베이스 파트너를 위해 사전 구축된 익스텐션 기능을 제공한다. 또한, 버텍스 AI를 통해 랭체인(LangChain)과 통합하고, 프라이빗 및 퍼블릭 API로 인증하며, 엔터프라이즈 보안, 개인정보 보호 및 규정 준수 제어를 통해 애플리케이션을 보호할 수 있다.

4) Google DeepMind, "Identifying AI-generated images with SynthID," Aug 29, 2023

디지털서비스 이슈리포트

- 챗봇 및 사용자 지정 검색 엔진과 같은 일반적인 차세대 AI 사용 사례를 빠르게 시작하려는 개발자를 위해 벡텍스 AI ‘검색 및 대화’ 기능은 인공지능 경험이 없어도, 그리고 대부분은 코드를 작성하지 않고도 시작할 수 있도록 도와준다.

듀엣 AI

올해 초, 구글은 클라우드 사용자가 더 효과적이고 효율적으로 작업을 수행할 수 있도록 파운데이션 모델에 기반한 상시 가동형 AI 협업 도구인 듀엣 AI를 소개했다. 이번에는 워크스페이스에서 듀엣 AI를 정식 버전으로 출시해 모든 사용자에게 글쓰기 도우미, 스프레드시트 전문가, 프로젝트 관리자, 회의용 메모 작성자, 창의적인 비주얼 디자이너를 제공한다.

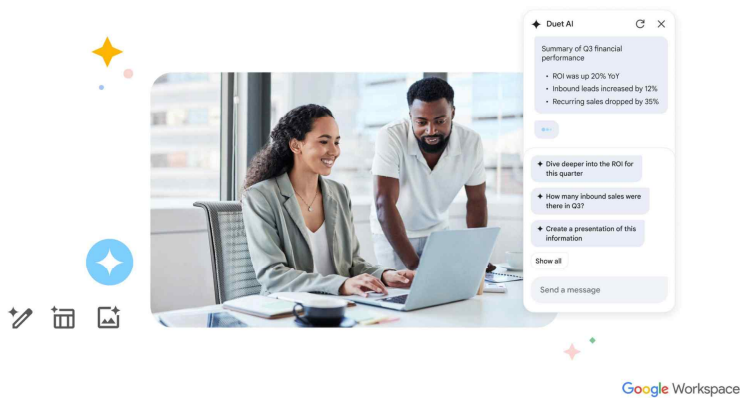


그림 3 워크스페이스에서 사용하는 듀엣 AI

워크스페이스는 현재 30억 명이 넘는 사용자와 천만 명 이상의 유료 사용자가 사용 중이다. 이번에 추가로 공개한 기능은 구글 미트(Meet)을 위한 기능과 구글 챗으로 위한 기능이다.

- 구글 미트에서 듀엣 AI: 듀엣 AI는 화상 통화 중에 메모를 작성하고, 회의 요약물을 전송하며, 18개 언어로 된 캡션을 자동으로 번역할 수도 있다. 또한 모든 회의 참가자가 명확하게 보고, 듣고, 이해할 수 있도록 스튜디오 룩, 스튜디오 조명 및 스튜디오 사운드를 발표했다.
- 구글 채팅에서 듀엣 AI: 듀엣 AI와 직접 채팅하여 콘텐츠에 대해 질문하고, 워크 스페이스에서 공유된 문서의 요약물을 확인하고, 놓친 대화를 따라잡을 수 있다. 또한 새로워진 사용자 인터페이스, 새로운 바로 가기, 향상된 검색 기능을 통해 대화를 놓치지 않고 파악할 수 있으며, 팀이 이미 공동 작업 중인 곳에서 회의를 시작할 수 있는 채팅 내 허들 기능도 제공한다.

디지털서비스 이슈리포트

이번 텍스트에서는 이제 듀엣 AI가 애플리케이션 개발, 운영, 데이터베이스 관리 및 마이그레이션, 데이터 분석 및 시각화, 사이버 보안을 지원하기 위해 구글 클라우드 전반에 걸쳐 세대 AI의 힘을 불어넣어 더 많은 영역에서 AI 지원을 제공할 수 있게 되었다고 발표했다. 전체 소프트웨어 개발 수명 주기에 걸쳐 전문가 지원을 제공하는데 여기에는 코드 생성, 소스 인용, 테스트 범위, API 설계 및 게시, 애플리케이션 마이그레이션 및 현대화 등을 포함한다. 전체 발표 내용을 요약하면 다음과 같다.

- 소프트웨어 개발: 전체 소프트웨어 개발 라이프사이클에 걸쳐 전문가 지원을 제공하여 개발자가 컨텍스트 전환을 최소화하여 생산성을 높일 수 있도록 지원함으로써 개발자가 흐름 상태를 더 오래 유지할 수 있도록 한다. 코드 완성 및 코드 생성 외에도 코드 리팩터링을 지원하여 애플리케이션을 더 빠르게 현대화할 수 있도록 도와주며, Apigee에서 듀엣 AI를 사용하면 모든 개발자가 간단한 자연어 프롬프트를 사용하여 API 및 통합을 쉽게 빌드할 수 있다.
- 애플리케이션 및 인프라 운영: 운영자는 구글 클라우드 콘솔에서 직접 여러 서비스에 걸쳐 듀엣 AI와 자연어로 채팅하여 인프라 구성, 배포 모범 사례, 비용 및 성능 최적화에 대한 전문가 권장 사항에 대한 '방법' 정보를 빠르게 검색할 수 있다.
- 데이터 분석: 빅쿼리에서 듀엣 AI는 SQL 쿼리 및 파이썬 코드 작성을 위한 상황에 맞는 지원을 제공하고, 전체 함수 및 코드 블록을 생성하고, 코드 완성을 자동 제안하며, 자연어로 SQL 문을 설명하고, 스키마 및 메타데이터를 기반으로 권장 사항을 생성할 수 있다. 이러한 기능을 통해 데이터 팀은 비즈니스를 위한 성과에 더 집중할 수 있다.
- 데이터베이스 가속화 및 현대화: 클라우드 스패너(Cloud Spanner), 엘로이DB 및 클라우드 SQL에서 듀엣 AI는 자연어를 사용해 데이터를 구조화, 수정 또는 쿼리하는 코드를 생성하는 데 도움을 준다. 또한 기존 번역 기술로는 변환할 수 없었던 저장 프로시저, 함수, 트리거, 패키지 등 데이터베이스 코드의 변환을 자동화하는 데 도움이 되는 데이터베이스 마이그레이션 서비스(DMS)에도 듀엣 AI를 사용할 수 있다.
- 보안 운영: 보안 전문가가 더 효율적으로 위협을 예방하고, 보안 워크플로우의 수고를 줄이며, 보안 인재의 수준을 높일 수 있도록 지원하는 크로니클 보안 운영, 맨디언트(Mandiant) 위협 인텔리전스 및 보안 커맨드 센터를 비롯한 보안 제품에 듀엣 AI를 도입하고 있다.

또한 GKE와 클라우드 런(Cloud Run)을 포함한 런타임에 듀엣 AI를 도입하도록 해서 클라우드에서 컨테이너를 실행하는 플랫폼 팀이 일상적으로 마주하는 수동적이고 반복적인 작업의 상당 부분을 줄일 수 있도록 했다. GKE 및 클라우드 런의 듀엣 AI는 프리뷰로 공개했다.

구글이 강조하는 면 중 하나는 사용자의 데이터가 곧 사용자의 데이터이고, 사용자의 코드, 듀엣 AI에

디지털서비스 이슈리포트

입력한 정보, 듀엣 AI에서 생성한 추천은 공유 모델을 학습하는 데 사용되거나 제품 개발에 사용되지 않는다는 점이다. 이게 구글의 원칙이라고 강조한다.

기타 발표

넥스트에서 발표한 추가 내용을 간략히 정리하면 다음과 같다.

- PaLM 2 모델의 컨텍스트 길이가 32k까지 확장되었다. 기존 8k 대비 4배가량 확장한 것이다.
- 코드 생성을 위한 코디(Codey) 모델이 지원하는 프로그래밍 언어 종류가 확대되었다. 베리로그를 포함해 약 30종 언어를 지원하기 때문에 언어 대부분을 포함한다고 볼 수 있다.
- 듀엣 AI 말고도 데이터 분석을 위한 추가 발표가 있다. 빅쿼리 스튜디오는 데이터 엔지니어링, 분석 및 예측 분석을 위한 단일 인터페이스이며, 앨로이DB AI는 표준 PostgreSQL보다 최대 10배 빠른 고성능 벡터 쿼리를 포함하여 생성형 AI 앱을 쉽게 구축할 수 있는 통합 기능 세트를 제공한다.
- 데이터 클라우드 파트너를 확대하고 있으며 애널리틱스 허브에 더 많은 데이터셋을 추가해서 고객이 생성형 AI를 학습하거나 구축하는 데 사용할 수 있다.
- 생태계를 더 확장하고 있으며 도큐사인과는 복잡한 계약서 및 기타 문서의 내용을 요약, 설명, 답변할 수 있는 스마트 컨트랙트 도우미를 생성하는 데 버텍스 AI를 어떻게 사용할 수 있는지 협력하고 있다. SAP와는 기업이 자동차 제조 간소화 또는 지속 가능성 개선과 같은 중요한 비즈니스 사용 사례에 차세대 AI를 적용하는 데 도움이 되는 SAP 데이터와 버텍스 AI를 활용한 새로운 솔루션을 구축할 수 있는지 확인하는 중이다.

02 클라우드와 오픈소스를 활용한 로스앤젤레스 카운티 투표 시스템 사례

| 윤대균 아주대학교 교수

선거 관리는 로스앤젤레스 카운티 등기등록국(LA County Registrar-Recorder)의 주요 업무 중 하나로 다른 무엇보다도 철통같은 보안을 요구하는 핵심 업무이다. 지난 선거에 대한 기록 및 앞으로 다가올 선거 관리뿐만 아니라, 유권자 등록/관리, 후보자 등록 및 정보 공유, 그리고 투표와 직결된 인적 물적 자원 관리를 책임짐으로써 투명하고 정확한 투표가 이루어질 수 있도록 하는 것이 핵심 미션이다. 특히 투표 운영은 다양한 공공기관의 미션 중에서도 단연 보안을 중시할 수밖에 없는 매우 민감한 업무라 할 수 있다.

최근 LA 카운티는 이처럼 보안을 최우선으로 하는 업무에 클라우드와 오픈소스를 적극 활용한 사례를 공개하였다.⁵⁾ 공공기관의 업무를 클라우드로 전환하는 과정에서 가장 애를 먹는 부분이 '보안'에 대한 불안감임을 감안할 때 투표와 같이 안전과 공공성을 가장 중요한 목표로 하는 업무에 클라우드를 적용한 것은 우리나라 공공기관 클라우드 전환 전략에도 참고할 만한 의미 있는 사례이다. 또한 오픈소스를 적극 활용한 것도 눈여겨볼 부분이다.

1. LA 카운티의 투표 시스템 - VSAP

본격적으로 클라우드 및 오픈소스 활용 사례를 소개하기 전에 LA 카운티에서 활용하는 투표 시스템에 대해 먼저 알아볼 필요가 있다. LA 카운티의 투표 시스템은 VSAP(Voting Solutions for All People)으로 대표된다. 이름에서 짐작할 수 있듯이 모든 유권자가 누구나 자신이 가장 편안하고 안전하게 투표를 행사할 수 있는 다양한 경험을 구현하도록 하는 것이 VSAP의 디자인 원칙이다. 이를 위해 2013년부터 잘 알려진 글로벌 디자인 회사인 아이디오(IDEO)와 함께 디자인 콘셉트를 도출하여 이를 VSAP이란 하나의 모델로 구현하였으며 지금도 계속 개선되는 중이다.

1.1 VSAP의 디자인 원칙

VSAP에서는 14가지 원칙을 채택하고 있다.⁶⁾ 여기에는 투표 시스템을 설계할 때 반드시 검토하여야 할

5) CIO.COM, "Los Angeles IT secures the vote with open source and the cloud", Aug. 11, 2023

디지털서비스 이슈리포트

중요한 요건이 잘 담겨 있다. 크게는 유권자를 위한 원칙, 공공 신뢰를 위한 원칙, 그리고 선거 운영 직원을 위한 원칙 이렇게 3가지로 분류하고 있다. 유권자를 위한 원칙으로는 다음과 같은 내용을 담고 있다.

- 투명성: 투표 관련 프로세스, 트랜잭션의 처리 등 제반 과정은 일반인이 쉽게 이해하고 확인할 수 있어야 한다. 특히 시스템 보안이나 가용성을 위태롭게 하지 않는 범위 내에서 관련 하드웨어 구성 요소 및 소스 코드를 검사할 수 있어야 한다.
- 접근성 및 프라이버시 : 어떤 유권자에게도 투표하는 일이 장벽이 되지 않고 참정권이 보장될 수 있도록 투표에 필요한 장소, 시간, 장비 활용 등 다양성이 허용되어야 한다. 또한, 장애를 가지고 있거나 영어 못하는 유권자도 언어 선택, 디스플레이 기기 조정, 오디오 투표나 특별한 기표 도구를 쓰는 등 대체 투표용지 방식을 통해 독립적으로 투표할 수 있어야 한다.

선거 운영 직원을 위한 원칙은 전반적인 투표 및 선거 이벤트 운영을 위해 필요한 제반 요건을 담고 있다. 이를 요약해 보면 다음과 같다.

- 확장성 및 유연성: 복잡한 투표용지 양식, 영어 이외의 다양한 언어, 매우 많은 수의 선거구, 지역구 및 지방 자치 단체와의 선거 통합을 수용할 수 있는 충분한 기술적 및 물리적 확장성을 제공해야 한다. 또한, 다양한 선거 유형이나 환경, 계속 변화하는 규제요건에도 안정성이나 보안성을 저해하지 않고 적용할 수 있는 '유연한' 기능을 제공해야 한다.
- 전력 및 네트워크 효율성: 시스템이 과도한 전력을 사용하거나 너무 복잡한 네트워크 연결로 인해 투표 시스템을 배치하는 데 어려움이 있어서는 안 된다.
- 프로그래밍 효율성: 운영 효율성과도 직결되는 것으로 투표소와 같은 현장에서의 시스템 부팅 및 프로그래밍 요구사항을 최소화함으로써 선거 당일 투표소의 선거 요원이 쉽게 작동할 수 있도록 설정할 수 있어야 한다.
- 이동성: 운반, 설치 및 효율적인 보관을 위해 충분히 가볍고 컴팩트해야 한다.

유권자와 선거 관리 직원의 경험을 강조한 원칙과 함께 시스템이 갖추어야 할 요건들도 명시되어 있다. '공공의 신뢰' 카테고리 분류되어있는 이 부분은 특히 성능, 보안, 그리고 비용과 직결된 부분이라 더욱 주목해 볼 필요가 있다.

- 신뢰(Trust): 투표 시스템에 대해 일반 대중이 신뢰할 수 있어야 한다. 투표자는 투표용지에 의도한 대로 기표했는지 반드시 확인할 수 있어야 하며, 추후 시스템에 대한 감사를 언제든지 수행할 수 있도록 모든 투표용지에 대한 확실한 기록을 생성할 수 있어야 한다. 즉, 유권자 및

6) <https://vsap.lavote.gov/principles/#1464310328742-c93e72f2-78ea>

디지털서비스 이슈리포트

후보자 그리고 모든 일반 대중에게 이 시스템을 통해 투표 결과가 제대로 집계된다는 사실을 입증할 수 있어야 한다.

- 무결성(Integrity): 시스템의 어떤 기능도 ‘사기(fraud)’ 의도에 의한 조작에 노출되지 않고 규정을 준수해야 한다. 투표수 및 시스템 성능을 모니터링하는 데 필요한 데이터를 생성하고 이에 대한 접근성을 높임으로써 감시 및 감사 기능을 제공해야 한다.
- 보안(Security): 선거 관리 직원 또는 관리인이 투표 장비를 안전하게 보관할 수 있는 잠금장치 및 보안 봉인과 같은 물리적 보안 조치 기능이 있어야 하며, 혹시라도 있을 수 있는 유무선 접속을 통한 비인가 접근 시도를 원천 차단할 수 있는 기능을 갖춰야 한다.
- 비용 효율(Cost Effective): 투표 시스템은 비용 효율적이어야 한다. 이러한 비용에는 조달, 운영 및 유지보수 비용 그리고 예상되는 시스템과 장비의 수명에 대한 고려도 필요하다.

1.2 VSAP 투표 시스템 개요 및 사용자 경험

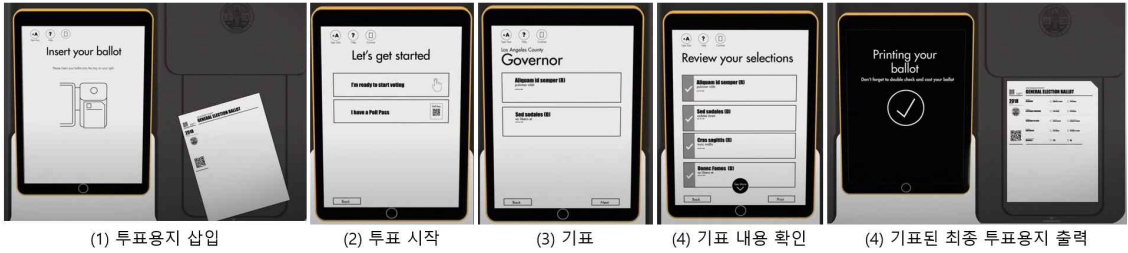
많은 실험과 인터뷰, 연구를 거쳐 나온 시스템의 핵심 결과물 중 하나인 투표기기를 보면 이러한 원칙이 어떻게 구현되었는지 확인해 볼 수 있다. (그림 4)



그림 4 VSAP 투표기

전체 투표 기기를 살펴보면 메인 화면, 시각 장애인을 위한 오디오 투표용 헤드폰, 조작 편의성을 더해 주는 키패드, 그리고 기표 용지 출력을 위한 프린터가 우측에 자리 잡고 있다. 다양한 언어 중 하나를 선택할 수 있을 뿐만 아니라 폰트 크기, 그리고 화면을 어두운 테마로 전환할 수 있도록 세부적인 부분까지 신경 쓴 것을 알 수 있다. 투표 과정은 지급된 투표용지를 단말기에 넣은 후 모든 기표를 마친 후, 자신의 기표 내용이 정확한지 확인한 후 투표용지를 출력하여 투표함에 넣는 것으로 마무리된다.(그림 5)

디지털서비스 이슈리포트



(1) 투표용지 삽입

(2) 투표 시작

(3) 기표

(4) 기표 내용 확인

(5) 기표된 최종 투표용지 출력

그림 5 VSAP 단말기에서의 투표 과정

투표소에서 위 투표 장치를 이용한 직접 투표 이외에도 다른 옵션도 제공된다. 우편을 이용한 투표도 지원하는데 직접 지정된 날짜에 투표소에 방문할 필요가 없기 때문에 많은 유권자가 선호하는 방식이기도 하다. 선거일 29일 전부터 시작되며 우편으로 송달된 투표용지에 기표한 후 다시 우편으로 보내거나, 혹은 가까운 투표용지 반납 장소 및 카운티 내 모든 투표센터에 제출하면 된다. 물론 투표용지는 영어 이외의 언어로도 요청할 수 있다. 또한, 투표용지를 수령할 때까지의 과정 및 기표 후 도착 및 개봉 여부까지 추적할 수 있는 서비스를 제공하며, 만일 수령한 투표용지가 파손되었거나, 또는 기표 실수를 했을 때는 새로 투표용지를 요청하여 받을 수 있다.

VSAP 투표 시스템의 흥미로운 또 다른 옵션은 이동식 투표센터이다. 앞서 소개한 VSAP 투표 장치를 카운티에서 공식적으로 운영하는 투표소 외에 둠으로써 접근성을 높이기 위함이다. 선거 하루 전에 설치하여 4시간에서 8시간 동안 운영할 수 있도록 하는 것인데 병원이나, 장애우를 위한 서비스 시설, 터미널과 같은 교통 시설 등에 설치 운영할 수 있다. 이를 위해 최소 공간, 카운티 차량을 위한 주차 시설 등 요건을 갖추어야 하지만 VSAP 장치를 활용하여 단시간에 설치할 수 있으며, 정규 투표소와 동일한 방식으로 안전하게 투표할 수 있는 것이 강점이다. 이동식 투표센터 정보도 일반 투표소와 마찬가지로 선거 관리국 사이트를 통해 확인할 수 있다.

2. VSAP 플랫폼 구축의 핵심 동력 – 오픈소스, 클라우드, 개방과 공유

민주주의에서 가장 중요한 참정 방식인 선거는 안정성, 무결성을 포함한 보안이 최우선이다. VSAP도 예외는 아니어서 유권자와 선거 운영자를 위한 최적의 경험을 제공하는 것과 함께 선거 시스템이 견고하고 안전하게 유지될 수 있도록 하는 것이 매우 중요하다. LA 카운티 등기등록국 CIO 아만 불라(Aman Bhullar)는 2020년 대통령 선거에 맞춰 오래된 구식 투표 인프라 개편을 주도하고, 특히 오픈소스 소프트웨어를 적극 활용하여 성공적인 혁신을 이룬 핵심 인물로 평가된다. AT&T와 사이버

디지털서비스 이슈리포트

보안 기술 분야에서의 협업, AWS 클라우드, 오픈소스 소프트웨어 등을 최대한 활용하여 20개 이상의 개별 시스템으로 구성된 VSAP 플랫폼 구축이 바로 그 혁신의 결과물이다.

2.1 오픈소스에 의한 오픈소스를 위한

불라는 2016년부터 LA 카운티 등기등록국 CIO로 재직하고 있다. 약 250명의 IT 전문가가 불라와 함께 팀을 이루어 LA 카운티의 투표 시스템을 구축하였는데, 긴밀하게 함께 협력한 주요 기업으로는 AT&T, 스마트매틱(Smartmatic), 아마존 AWS를 들 수 있다. AT&T와는 사이버보안 분야, 스마트매틱은 투표기 제작, 그리고 아마존 AWS 클라우드 서비스를 핵심 인프라로 활용하였다. 이 팀은 리눅스 배포의 한 버전인 센토스(CentOS)에서 실행되는 많은 애플리케이션을 개발하여 VSAP 플랫폼을 구축하였으며, 이렇게 개발한 애플리케이션은 오픈소스로 공개하여 누구나 이를 활용할 수 있도록 했다. 불라는 이러한 개방과 공유는 VSAP 개발의 핵심 원칙이라고 강조하며 이미 오픈소스로 공개된 70%의 코드 이외의 나머지 30%의 독점(Proprietary) 코드도 오픈소스로 전환하여 공개적으로 활용할 수 있도록 캘리포니아주와 협력하여 '규제 프레임워크'를 구축하고 있다고 설명했다. 이러한 활동을 통한 성과로 불라와 그의 팀은 IT 혁신 및 리더십 부문에서 2023 CIO 어워드를 받는 등 많은 상을 받았다.

LA 카운티 투표 시스템에서 가장 자신 있게 내세우는 부분은 투표 기계와 애플리케이션을 자체 개발한 것이다. 앞서 소개한 투표 장치는 온라인과 오프라인 경험을 완벽하게 통합하는 것이 핵심 특징이라고 볼 수 있는데, 예를 들어 투표자의 입력을 '온라인'으로 기록함과 동시에 기계와 인간이 모두 판독할 수 있는 '투표용지'로 변환함으로써 온라인 저장과 오프라인 기록을 모두 확보할 수 있게 한 것이다. 이는 투표의 무결성 및 향후 사후 검증을 위해서도 활용될 수 있다. 오프라인으로 출력된 투표용지와 함께 온라인 버전 투표용지는 아파치 카산드라 데이터베이스에 법률로 정한 기간 보존된다. 카산드라 데이터베이스는 대표적인 NoSQL 데이터베이스로 확장성 및 유연성이 뛰어나고 빅데이터 분석에도 용이하다. 또한, 우편 투표 또는 투표 센터를 통해 받은 직접 투표용지를 스캔하여 집계하는 애플리케이션을 개발하여 오픈소스로 공개했다. 이 애플리케이션은 센토스OS와 카산드라 DB에 기반하여 외부 네트워크에 연결될 필요 없이 투표자 데이터베이스에 접근할 수 있다. 이 밖에도 새로운 투표 시스템 테스트 및 교육을 위한 샘플 투표용지 애플리케이션도 개발/공개하였다.

2.2 AWS 기반의 클라우드 전환

클라우드 전환은 아마존 AWS를 중심으로 진행하였다. 유권자 대면 서비스나 투표 정보 제공 등의

디지털서비스 이슈리포트

서비스는 모두 AWS 클라우드 기반으로 구현되었다. AWS 서비스는 특히 팬데믹 기간 내내 LA 카운티의 필수 요소가 되었다. 예를 들어, LA 카운티 공공서비스국, 아동가족서비스국 등 카운티의 40개 정부 기관 중 몇몇은 팬데믹 기간동안 원격 근무를 촉진하기 위해 아마존 앱스트림(AppStream)을 사용했다. 이렇게 AWS 클라우드의 활용이 늘어남에 따라, 2020년 대통령 선거를 앞두고 불라와 그의 팀은 유권자 지원 및 선거구 서비스를 제공하기 위해 기존에 운영하던 .NET 기반 웹사이트도 VSAP 투표 시스템 도입과 함께 AWS로 마이그레이션하였다.

클라우드 전환을 통해 카운티의 선거 인프라는 더욱 확장이 용이하고 안전해졌으며, 투표소 위치, 투표결과 확인 기능, 장애인이 이동 투표 차량과 같은 지원을 요청할 수 있는 링크 제공 등 시민들이 투표하는 데 필요한 모든 정보에 액세스할 수 있게 되었다.

2.3 개방과 공유, 협업을 통한 사이버 보안 프로그램

클라우드 전환에도 불구하고 안전하고 투명한 ‘투표행위’와 관련된 제반 업무는 클라우드와 무관한 수동 프로세스로 남아있을 수밖에 없음을 불라(Bhullar)는 강조한다. 기표가 끝난 투표용지는 한번 스캔 후 일반 상자에 보관되어 철저히 보호되어야 한다. 즉, 실제 투표 경험은 클라우드 기반이 아니며 외부 세계와는 격리되어 네트워크에도 연결되어 있지 않다. 투표 기간 내내 관련 장치는 네트워크에 연결되지 않는다. 투표 기간과 관계없이 선거에 중대한 악영향을 끼치는 잘못된 정보와 허위 정보로부터 보호하기 위한 강력한 사이버 보안 프로그램은 여전히 필요하다.

VSAP 투표 시스템 및 전반적인 선거 관리를 위한 보안 프로그램은 AT&T 사이버시큐리티와 협력하여 개발했다. CSOC(Cyber Security Operation Center)와 같은 핵심 기능은 투표 기간 중 상시 24간 운용되며 잠재적 보안 위협 및 해킹 시도를 모니터링한다. 이를 위해 기존 온-프레미스 환경뿐만 아니라 클라우드 환경에서 모든 보안 관련 데이터를 실시간으로 수집/분석함으로써 위협에 대한 조기 감지 및 대응을 할 수 있도록 했다. 최고 수준의 보안을 달성하기 위해 EI-ISAC(Elections Infrastructure Information Sharing & Analysis Center), FBI, CISA(Cybersecurity & Infrastructure Security Agency) 그리고 OTX(Open Threat Exchange)의 데이터도 활용한다. OTX는 전 세계 사이버 보안 연구자와 화이트 해커 등 관련 분야의 전문가 커뮤니티가 활동하는 플랫폼이다. 연구 결과나 새로 발견된 보안 위협 정보를 공유하며 협업할 수 있는 공간이다. VSAP의 사이버 보안 프로그램에는 제로 트러스트 모델에 입각한 신원 및 액세스 관리, 엔드포인트 모니터링과 같은 기능도 포함되어 있다. 이는 바이든 정부 차원에서 추진하는 사이버 보안 고도화 전략과도 일맥상통한다.

디지털서비스 이슈리포트

3. 시사점

투표 시스템과 같이 민감한 공공 서비스도 오픈소스와 클라우드를 활용하여 혁신을 만들어 낼 수 있다는 것이 이 글의 핵심 주제다. 따라서 오픈소스의 중요성 및 민간 클라우드 활용의 당위성 등에 대해서는 굳이 다시 언급할 필요는 없겠다. 다만 필자의 관점에서 특히 주목하고 싶은 것은 VSAP 개발을 주도한 조직의 역할이다.

개발과 운영을 책임진 조직의 전문성에 대해 언급하지 않을 수 없다. LA 카운티 등기등록국 CIO 아만 불라(Aman Bhullar)와 그의 팀이 VSAP 전체 개발 과정에 깊이 관여했음을 여러 사례를 통해 알 수 있다. 600만 인구 LA 카운티의 한 조직에서 투표 시스템을 개발하기 위해 250명의 IT 전문가가 참여했다는 사실이 이를 뒷받침하고 있는데, 실제 VSAP 애플리케이션을 자체 개발하고 이를 오픈소스로 공개하기 위해서는 우리나라 공공기관의 전형적인 하도급 개발 방식으로는 거의 불가능할 것이다. 스마트매틱과 협업을 통해 투표 기기 개발을 했다고 하지만 일관성 있는 사용자 경험 및 VSAP 보안 요구사항 준수를 위해서는 CIO조직이 처음부터 끝까지 주도할 수밖에 없다. 이렇게 어느 정도 규모가 되는 전문 IT 조직을 갖추고 직접 프로젝트를 수행하기에 혁신을 이끌 수 있었다. 디지털 플랫폼 정부 관점에서 '민간 주도 혁신'은 매우 중요한 키워드이지만 정부 조직 내 전문성을 바탕으로 플랫폼을 설계하고 운영할 수 있는 역량이 없다면 공염불에 불과한 것이다. 즉 민간 주도 혁신을 논하기 전에 이를 뒷받침하는 공공 정부 조직에서의 전문성이 담보되어야 한다는 사실이다. LA 카운티의 VSAP은 정부 IT 조직의 전문성이 얼마나 중요한지를 보여주는 단적인 예다.

사이버 보안 프로그램 개발 과정에서는 정부 내 많은 부처로부터의 데이터 확보가 눈에 띈다. LA 카운티 당사자가 아닌 다른 부처에서 민감한 데이터를 자발적으로 제공하지는 않았을 것이다. 해당 프로젝트를 주도하는 CIO 조직이 나서서 기획하고 조율하지 않으면 한 발짝도 나가기 힘들었을 것으로 쉽게 상상할 수 있다. 여기서도 CIO 조직의 전문성이 발휘되었을 것으로 쉽게 짐작해 볼 수 있다. 이들이 더 나아가 OTX와 같은 개방형 플랫폼을 활용한 것은 향후 발전을 위한 탁월한 선택이었다고 본다. 또한, 당장 기존 규제에 묶여 오픈소스로 개방하지 못하는 독점(Proprietary) 코드도 점차 오픈소스화하기 위해 상위 기관인 캘리포니아 주와 협력을 하는 부분도 이들이 개방과 공유의 철학을 진심으로 실천한다는 것을 엿볼 수 있는 부분이다.

끝으로 2013년 처음 프로젝트 시작부터 지금까지 모든 활동이 꾸준히 잘 기록되고 있다는 사실을 짚어보고 싶다. VSAP 홈페이지⁷⁾에는 VSAP에 대한 소개뿐만 아니라 그간의 여정이 상세히 소개되어

디지털서비스 이슈리포트

있다. 여기에는 VSAP 철학, 서비스로서의 목표와 비전, 그리고 이런 것을 설계와 개발에 어떻게 담았는지를 찾아볼 수 있다. 성공한 프로젝트 결과로써의 끝이 아닌 미래에 대한 새로운 미션 또한 제시하고 있는 것이다. 일관된 원칙과 철학을 바탕으로 장기적으로 진화와 혁신이 반복되는 과정을 거치다 보면 자연스럽게 미래에 대한 모습도 그려지게 마련이다. 우리나라 IT 프로젝트에서 “왜 하는가?”라는 질문에 대한 답이 궁색한 이유가 이런 ‘기록’의 부재에서 시작된다. 이것이 해결된다면 더욱 뛰어난 결과물을 제시할 수 있을 것이다.

7) <https://vsap.lavote.gov/>

03 맥킨지 리포트: 2023년 중요 기술 동향과 전망

| Senior Program Manager 김영욱

들어가며

액센츄어 리포트⁸⁾에서는 “기술의 성숙도는 데이터와 해당 기술뿐만 아니라 조직의 전략, 인재, 문화 등 핵심 역량을 적절히 조합하여 완성형 모델로 만드는 데 달려 있다”라고 한다. 기업과 조직은 비즈니스 발전을 촉진할 수 있는 신기술 잠재력에 대한 투자와 인재 확보에 있어 어느 때보다도 높은 관심의 시간을 보내고 있다. 생성형 AI는 이러한 기술 르네상스를 연 공로가 탁월하지만, 이는 곧 다가올 많은 기술 발전 중 하나에 불과할 것이다. 이러한 새로운 기술들은 지속 가능하고 포용적인 성장을 주도하고 복잡한 글로벌 과제를 해결할 수 있을 것으로 기대한다.

매년 새로운 최신 기술 동향을 발표하는 맥킨지에서 2023년 가장 중요한 기술 동향을 분석하고 전망한 결과 리포트⁹⁾를 발표했다. 그 기술들이 아직 도입 및 확장 초기 단계에 있지만, 이 분석 결과를 통해 잠재적인 사용 사례를 이해하고 미래 기회 가치를 선점하기 위한 인재 채용과 핵심 기술을 파악하는 것은 기업의 생존에 있어 매우 중요하다.

이 리포트는 모두 15개의 기술을 제시하였고, 관심도, 혁신성, 투자에 대한 정량적 지표를 조사하여 각 기술의 모멘텀을 측정하고 각 기술 동향의 장기적인 특성과 상호 의존성을 고려하여 주변 기반 기술, 불확실성, 의문점 등을 심층적으로 서술하였다. 특히 올해 보고서의 특징으로는 결과 분석에 ‘인재(talent)’라는 중요한 차원을 새롭게 추가했다는 점이다. 예를 들어 이 인재 차원은 각 트렌드와 가장 관련성이 높은 직무에 대한 인재 수급에 대한 데이터를 제공하여 흥미로운 분석 결과를 보여준다.

아래 그림 6의 경우처럼, 채용공고는 인공지능, 차세대 소프트웨어 개발, 클라우드와 같은 세 개의 축이 이끌어 가고 있는 상황이다.

8) Accenture, “The art of AI maturity”, 2023

9) McKinsey Digital, “McKinsey Technology Trends Outlook 2023”, Jul 20, 2023

디지털서비스 이슈리포트

Tech trend job postings,¹ 2021-22, thousands

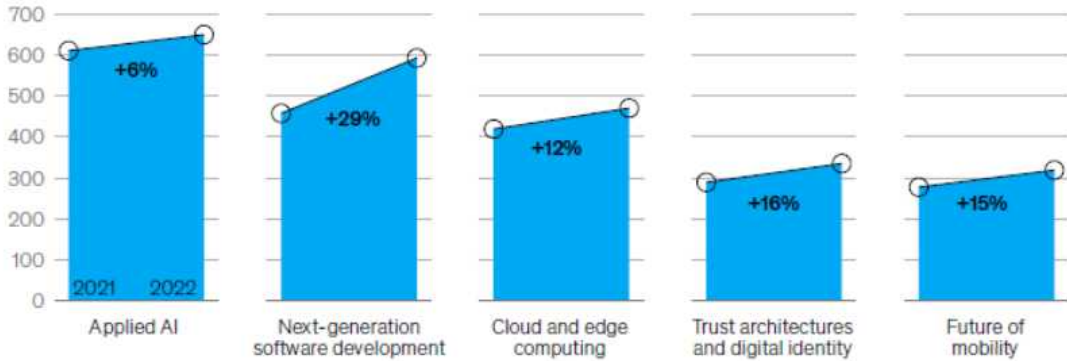


그림 6 2021-22 기술직군 채용공고 동향 (출처: 맥킨지)

15개의 기술 동향 중 클라우드 컴퓨팅과 인공지능 기술을 포함 가장 무게감 있는 5가지를 골라 집중적으로 정리해 본다.

2023년 5가지 핵심 기술

1) 응용 AI (Applied AI)

동향

모든 산업 분야의 기업은 머신러닝, 컴퓨터 비전, 자연어 처리와 같은 인공지능 기능을 통해 데이터를 사용하고 인사이트를 도출, 프로세스를 자동화하고, 사용자 기능을 추가 또는 확장하여 더 나은 의사 결정을 내릴 수 있다. 응용 AI를 통해 얻을 수 있는 잠재적 경제적 가치는 작년 17조 달러에서 26조 달러로 증가했고 조직의 AI 도입 비율은 2017년 20%에서 2022년 50%로 두 배 이상 늘었다. 그러나 기업이 기술의 잠재력을 최대한 활용하기 위해서는 조직, 기술, 윤리 및 규제 문제의 해결이 필요하다.

- 역량 강화를 위한 투자** : 2021년 1,468억 달러로 최고치를 기록한 후 2022년 1,040억 달러로 감소했지만, 다른 기술과 비교하면 평균 수준보다 앞서는 속도를 보인다. 모델 정확도 등의 분야에서 지속적인 개선이 보인다. 예를 들어, 이미지 분류 시스템을 훈련하는 데 드는 비용은 2018년 이후 63.6% 감소했으며, 훈련 시간은 94.4% 개선되었다.¹⁰⁾ 새롭게 부상하는 기술과

10) Stanford AI Index, “Artificial Intelligence Index Report 2022”, Mar 16, 2022

디지털서비스 이슈리포트

결합하면 응용 AI의 잠재력을 더욱 높일 수 있는데 예를 들어, 생성형 AI의 기반이 되는 기초 모델은 메모 및 로그와 같은 대량의 비정형 데이터를 처리하여 성능을 최적화하는 작업에 큰 도움을 줄 수 있다.

- 2. 규제 조치 가속화** : 기술이 발전함에 따라 데이터의 사용 및 결정에 대한 오용 가능성도 커진다. 인공지능의 윤리적 오용과 관련된 사건을 추적하는 AIAAIC(AI, 알고리즘 및 자동화 사고 및 논란) 리포지토리에 따르면 알고리즘 및 자동화와 관련된 윤리적 이슈가 2012년 이후 26배나 증가했다고 한다. 민간 기업의 인공지능 개발 가속화에 힘입어 기초적인 AI 모델을 규제하는 유럽연합의 AI 법이 의회 위원회 승인을 거쳐 법제화를 앞두고 있다.
- 3. 기술 도입은 완료 상태** : 전 세계적으로 인공지능 채택률은 2017년의 두 배 이상이지만, 최근 몇 년 동안 AI를 사용하는 조직의 비율은 약 50~60%로 평준화되었다. 이미 인공지능 기술을 도입한 기업은 자연어 생성이나 컴퓨터 비전과 같이 사용하는 기능의 수를 두 배 가까이 늘릴 계획이다.¹¹⁾

불확실성

인공 기술 적용에 영향을 미치는 주요 불확실성은 다음과 같다.

- 머신러닝 산업화를 위한 솔루션과 IT 인프라의 기술적 발전에도 불구하고 인재 및 투자금과 같은 가용 리소스가 부족하면 AI 애플리케이션 배포 계획에 영향을 미칠 수 있다.
- 2022년 주요 리스크로 사이버 보안과 개인정보 보호, 데이터 위험과 취약성뿐만 아니라, 신기술에 대한 규제 및 규정 준수는 인공지능 연구 및 애플리케이션에 영향을 미칠 수 있다.
- 데이터 거버넌스, 형평성, 공정성을 포함한 윤리적 고려사항은 인공지능 제품 사용에 영향을 미칠 수 있다

2) 머신 러닝 운영(MLOps)

MLOps는 기업에서 머신러닝 애플리케이션을 확장하고 유지하는 데 필요한 엔지니어링 실행을 말한다. 이것은 기업 내 기술 도구 에코시스템에 의해 활성화되고 지원된다. MLOps 도구는 기업이 파일럿 프로젝트 수준에서 실행할 수 있는 비즈니스 제품으로 전환하고 분석의 규모를 빠르게 확장하는 데 도움을 주고, 솔루션 및 프로덕션에서 발생하는 문제를 빠르게 선제적으로 식별 및 해결하고 팀의 생산성을 개선하는 데 도움을 준다.

11) McKinsey, "The state of AI in 2022—and a half decade in review", Dec 6, 2022

디지털서비스 이슈리포트

동향

- 투자 증가** : MLOps 솔루션을 제공하는 기업에 대한 투자는 2022년에도 누적 투자액이 47억 달러로 강세를 유지했다. IDC는 2024년까지 60%의 기업이 MLOps를 도입하고 구현할 것으로 예측한다.¹²⁾ 인공지능 기술을 통해 더 높은 수익을 창출하는 기업이 MLOps를 실행할 가능성이 더 높다는 사실인 만큼 이 분야에 대한 투자는 꾸준히 증가할 것으로 예상된다.
- 생태계는 빠르게 진화** : 2022년은 중요한 인수, 합병, 파트너십이 있었다. 알테어(Altair)가 래피드마이너(RapidMiner)를 인수했고, 스노우플레이크(Snowflake)는 미스트 AI(Myst AI)를, 휴렛팩커드가 파키덤(Pachyderm)을 인수했습니다. 데이터브릭스(Databricks)는 인기 높은 오픈 소스 MLOps 프레임워크의 v 2.0을 발표했다. 갠트리(Gantry)는 실시간 ML 모델 생산 성능과 이를 최적화하는 방법을 결정하는 플랫폼을 출시했다. MLOps를 고민하는 기업들에겐 선택의 폭이 넓어진 생태계가 완성되었다.

불확실성

기업의 의사 결정권자는 MLOps를 도입하고 실행할 때 다음 사항에 관해 충분한 고려가 필요하다.

- 기업과 조직에 이미 정형화된 업무 프로세스에 특화된 MLOps를 구축하려면 선행 투자와 리소스가 필요하다.
- 산업 규모에서 ML 솔루션을 유지 관리하려면 명확하고 동의된 프로세스와 책임이 중요하다.
- 빠르게 진화하는 시장에서 벤더 종속을 피하고 기존 벤더의 에코시스템 밖에 있는 다른 벤더가 제공하는 새로운 제품에서 가치를 실현할 수 있어야 한다.
- 조직이 사용 사례 요구에 맞는 수준과 적합한 솔루션에 투자하고 있는지 확인하여 잘못된 가능성을 미연에 방지해야 한다.

3) 생성형 AI

생성형 AI는 인공지능에서 가장 주목받는 기술이다. 이전의 AI와 달리 유사한 형식의 비정형 데이터에서 학습한 정보를 기반으로 텍스트, 오디오, 비디오, 이미지, 코드, 시뮬레이션, 심지어 단백질 서열이나 소비자 여정과 같은 새로운 비정형 콘텐츠를 생성할 수 있다. 또한 핵심 기술인 기초 모델은 요약, 분류, 초안 작성 등 다양한 작업에 적용될 수 있다. 이전 세대의 AI 모델은 한 가지 작업만 수행할 수 있는 '협소성'을 가진 경우가 대부분이었다.

12) CIO.com, "Making the most of MLOps", May 26, 2022

디지털서비스 이슈리포트

비즈니스 맥락에서 생성형 AI는 새로운 사용 사례를 발굴할 수 있을 뿐만 아니라 기존 사용 사례의 속도, 확장 또는 개선도 가능하다. 생성형 AI의 비즈니스 잠재력과 가치 사슬은 새로운 제품 개발, 고객 경험 향상에 기여할 것으로 예상되나 가장 큰 영향은 직원 생산성 및 경험 개선에 있을 것으로 예상된다.

초기에는 많은 업계에서 생성형 AI를 주로 업무 초안을 작성하거나, 가설을 생성하거나, 전문가가 작업을 더 빠르고 더 잘 수행할 수 있도록 지원하는 보조 기술로 사용하게 되는데, 여기에는 다음과 같은 두 가지 공통점이 있다.

- 애플리케이션에서 생성된 부정확한 콘텐츠와 지적 재산권 문제 등 결과물을 확인할 수 있는 전문가가 전체 과정에 있다.
- 기존 워크플로에 사용되므로 채택 및 변경 관리가 용이하다

조직이 생성형 AI 기반 애플리케이션을 보조적인 수준에서 위험도가 높은 고효율 사용 사례를 지원하는 완전 자동화로 발전시키기까지는 약간의 시간이 걸릴 것으로 예상된다.

동향

1. **파일럿 투자의 증가** : 생성형 AI에 대한 투자는 2020년에 비해 425% 증가했고, 현재 AI 연구의 80%가 생성형 AI에 초점을 맞추고 있기 때문에 금융 서비스에서 생명 과학에 이르는 다양한 산업 분야의 기업들이 엔터프라이즈 사용 사례를 실험하기 시작한 단계에 접어들었다. 코히어(Cohere)와 같은 업체는 대기업이 원하는 높은 수준의 지적 재산권 보호, 소비자 개인정보 보호, 저렴한 비용의 LLM을 자사 환경에 제공하고 있다. 많은 스타트업들이 LLM 플랫폼을 기반으로 본인의 서비스를 구축하거나 오픈 소스 모델을 확장하고 있는 과정에 구글은 검색 생성형 경험(Search Generative Experience)과 바드 챗봇을 비롯한 여러 구글 제품을 구동할 새로운 LLM인 PaLM 2와 같은 기반 기능을 발표했다. 오라클, SAP, 세일즈포스와 같은 전통적인 비즈니스 소프트웨어 벤더는 기존 업무용 제품에 생성형 AI 기능을 통합하기 위해 상당한 투자를 하고 있다.
2. **GPT4와 그 이후** : GPT-3가 변호사 시험 응시자 중 하위 10%의 성적을 기록한 반면, GPT-4는 상위 10%의 성적을 기록했다. GPT-4는 이미지와 텍스트를 입력으로 사용할 수 있고, 최대 25,000 단어(GPT-3의 경우 4,000 단어)까지 처리할 수 있으며, 정확한 응답을 생성할 가능성이 40% 더 높다. 멀티모달 입력 사용 및 오케스트레이션과 같은 여러 가지 복잡한 애플리케이션이 나타나고, AutoGPT 및 BabyAGI와 같은 애플리케이션을 통해 새로운 레시피를 고안하는 등의 작업 수행이 가능하다.
3. **하드웨어 가속기 설계 활동** : 구글은 이전 버전에 비해 시스템 성능이 약 10배 향상된 4세대 텐서

디지털서비스 이슈리포트

프로세싱 유닛(TPU v4)을 개발했다.¹³⁾이런 경향은 애플의 뉴럴 엔진과 같이 소비자 하드웨어 및 엣지 시스템에서도 LLM을 보다 효율적으로 실행할 수 있는 특정 설계를 갖춘 제품들이 출시되고 있다.¹⁴⁾

4. 생성형 AI는 2018년부터 인재 수요가 빠르게 증가하고 있다. 2023년의 채용 수요는 관심과 투자 증가로 인해 훨씬 더 높는데, 이 분야의 채용 증가는 다른 기술 트렌드와는 다른 경향을 보인다. 규제 담당 관리자급 채용공고가 많은 것은 기업이 생성형 AI 전략 개발이 빠르게 필요하고 그 전략을 통해 다음 행보를 결정하고자 함을 나타낸다. 개발자나 엔지니어의 채용률이 상대적으로 낮다는 것은 내부 역량 구축에 투자하기 전에 사용자 친화적인 오픈소스 모델을 시범 운영하여 먼저 가치를 검증하려는 의도를 보여준다고도 해석할 수 있다.

5.

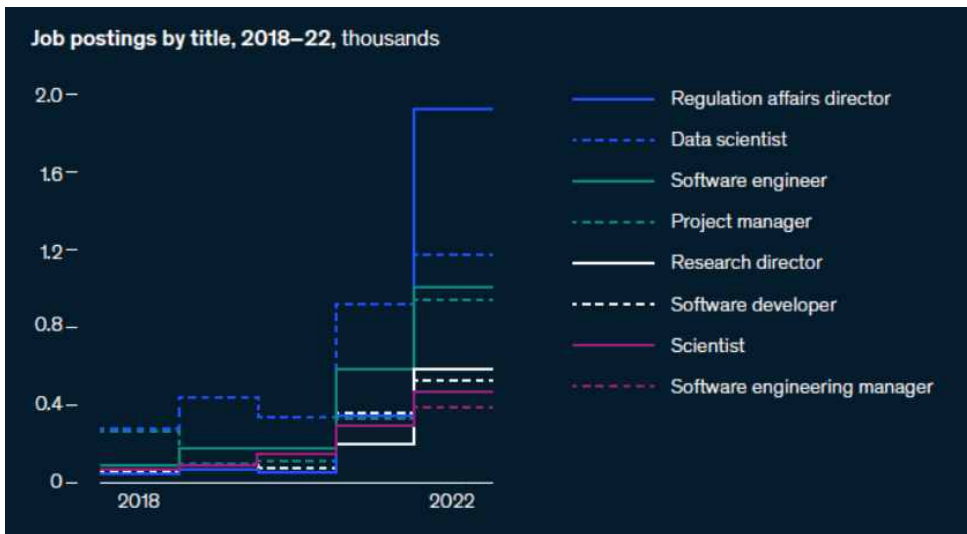


그림 7 생성형 AI 채용공고 동향 (출처: 맥킨지)

불확실성

생성형 AI에 영향을 미치는 주요 불확실성은 다음과 같다.

1. 고객 및 보호 대상의 데이터 유출 위험 및 취약성과 관련한 사이버 보안 및 개인정보 보호 문제가 있다.
2. 데이터 거버넌스, 정의와 공정성, 책임성, 설명 가능성 등 생성형 AI의 책임감 있는 사용과 관련된 윤리적 고려사항이 있다.

13) Google Cloud, “Google’s Cloud TPU v4 provides exaFLOPS-scale ML with industry-leading efficiency”, Apr 6, 2023

14) Apple, “Deploying Transformers on the Apple Neural Engine”, Jun 2022

디지털서비스 이슈리포트

3. 규제 및 규정 준수는 생성형 AI와 그 잠재적 응용 분야에 관한 연구에 영향을 미칠 수 있다.
4. 오픈 소스 모델로 생성된 콘텐츠의 저작권과 소유권 및 보호는 아직 정의되지 않은 문제다.
5. 학습 모델이 기하급수적으로 더 많은 컴퓨팅 리소스를 사용함에 따라 환경에 미치는 영향이 증가한다.

4) 클라우드/엣지 컴퓨팅

앞으로 기업은 조직 내 온프레미스 환경부터 가까운 위치 (엣지), 소규모 지역 데이터 센터, 원격 하이퍼스케일 데이터 센터에 이르기까지 여러 위치 지점에 컴퓨팅과 스토리지가 포함된 인프라 공간을 활용하게 될 것이다. 엣지 컴퓨팅은 원본에 더 가까운 곳에서 데이터를 더 빠르게 처리할 수 있는 유연성(초저지연: Ultra-low latency)을 제공하며, 클라우드에 비해 데이터 주권을 확보하고 데이터 프라이버시를 강화할 수 있어 다양한 새로운 사용 사례를 창출할 수 있다. 최종 사용자에게 목적지까지의 거리 단축의 의미는 데이터 전송 지연과 비용을 줄일 수 있을 뿐만 아니라 더 많은 관련 데이터 집합에 더 빠르게 액세스할 수 있어 기업이 데이터 보존 관련 법률을 준수하는 데 도움이 된다.

이런 경향에도 불구하고 퍼블릭 클라우드는 기업에서 지속적으로 중요한 역할을 할 것이다. 시간에 민감하지 않은 컴퓨팅 사용 사례를 훨씬 더 나은 규모의 경제로 수행하여 기본 비즈니스의 안정성을 구현할 수 있다. 클라우드와 엣지 리소스의 유연한 통합을 통해 사용자는 클라우드의 혁신, 속도, 민첩성을 엣지까지 확장할 수 있다. 실시간 시스템을 구축하여 혁신을 가속화하고 생산성을 높이며 비즈니스 가치를 창출한다.

동향

1. **마이그레이션 속도 둔화¹⁵⁾** : 기업들이 퍼블릭 클라우드로 마이그레이션하고 있지만, 데이터 프라이버시 및 지연 시간 관련 문제와 치솟는 비용관리의 어려움으로 인해 마이그레이션 속도가 늦어지고 있다. 업타임 인스티튜트 글로벌 데이터 센터의 연구 결과에 따르면 일부 기업의 경우 클라우드에서 데이터 센터로 '회기 이전'한 경험이 있다고 한다. 그러나 클라우드로 이전한 기업 중 클라우드를 완전히 포기한 비율은 6%에 불과하다. 대다수는 온프레미스 및 퍼블릭 클라우드 옵션을 사용하는 하이브리드 접근 방식을 사용한다.
2. **엣지 컴퓨팅의 지속적 확산** : 다양한 산업 분야에서 400개 이상의 엣지 사용 사례가 확인되었으며

15) The Register, "Oh dear, AWS. Cloud growth slowing as customers get a dose of cost reality", Jan 19, 2023

디지털서비스 이슈리포트

향후 5년간 전 세계적으로 두 자릿수 성장을 보일 것으로 예상된다.¹⁶⁾ 원격 공장과 같은 특정 위치에서의 안정적인 네트워크의 부족 현상을 비롯한 다양한 이유로 엣지 도입이 증가하고 있다. 엣지 컴퓨팅을 통해 데이터가 생성된 곳에서 데이터를 직접 처리함으로써 더 빠른 속도와 정확성, 효율적인 데이터 분석을 통해 올바른 비즈니스 의사 결정이 가능해진다.

3. 클라우드에서 고성능 컴퓨팅과 AI/머신러닝 사용이 증가 : 워크로드의 증가를 활용하고 하드웨어를 최적화하기 위해 클라우드 제공업체는 여러 기술 파트너십을 강화할 뿐만 아니라 구글의 텐서 처리 장치(TPU) 및 AWS의 니트로(Nitro) 시스템과 같이 기업 내 칩 설계에도 투자하고 있다.
4. 인재 수급 상황에서 보면 리눅스, 파이썬, 클라우드 컴퓨팅과 같은 기본적인 영역에서는 이미 포화 상태를 보인다. 도입 시기가 지나고, 본격적인 운영 시기가 되면서 하이퍼스케일러 운영 및 데브옵스와 인프라 관리 부문 인재들의 수요가 증가하고 있다.



그림 8 클라우드 컴퓨팅 인재 가용성 (출처: 맥킨지)

불확실성

클라우드 및 엣지 컴퓨팅에 영향을 미치는 주요 불확실성은 다음과 같다.

1. 엣지 컴퓨팅은 기존 클라우드 컴퓨팅의 규모의 경제를 활용하지 못하기 때문에 엣지 노드 및 디바이스 수가 증가함에 따라 확장에 따른 여러 문제가 발생할 수 있다.
2. ROI 가시성 부족, 엣지 개발에 대한 전반적으로 더 긴 수익 경로가 예상된다.
3. 부가가치 사용 사례에 대한 고객의 이해 부족, 파일럿에서 대규모 구현으로 확장하기 위한 대규모 투자 요구사항, 엣지 컴퓨팅 위치의 단기 공급 과잉, 복잡한 기술 스택 요구사항, 즉시 배포할 수 있는 솔루션의 부족 등이 불확실성이 있다.

16) IDC, "New IDC Spending Guide Forecasts Edge Computing Investments Will Reach \$208 Billion in 2023", Feb 15, 2023

디지털서비스 이슈리포트

5) 몰입형 실감 현실 기술 (Immersive-Reality technologies)

몰입형 현실 기술은 공간 컴퓨팅을 사용하여 물리적 공간을 해석하고, 실제 환경에 데이터, 사물, 사람을 추가하여 시뮬레이션하며, 증강 현실(AR), 가상 현실(VR), 혼합 현실(MR)이 제공하는 다양한 수준의 몰입감으로 가상 세계에서 상호작용을 가능하게 한다. 2022년에는 AR과 VR에 대한 총투자액이 전 해에 비해 감소했지만, 이 분야는 기술 투자에 관심이 높은 것은 분명하다. 퓨어스토리지의 연구에 따르면 새로운 메타버스는 2030년까지 소비자와 기업 사용 사례에서 최대 5조 달러의 가치를 창출할 수 있을 것으로 예상된다.¹⁷⁾

동향

- VR 및 AR 기기에 대한 투자와 소비자 채택이 감소** : 평면 기기에서의 AR 채택은 증가했다. 몰입형 현실 기술에 대한 투자는 2018년 140억 달러, 2021년 320억 달러, 2022년 160억 달러로 전 세계 VR 헤드셋 및 AR 기기 출하량이 12% 이상 감소했다. 그러나 스마트폰과 같은 평면 디바이스에서의 AR 사용은 계속 증가하고 있다. AR 앱 설치 기반은 2016년 10억 3천만 개에서 2022년 60억 6천만 개로 약 6배가량 증가했다.
- 새로운 기기 출시** : 하드웨어 혁신은 주춤했지만, 2022년에는 새로운 VR 헤드셋이 출시되었다. 지속적으로 새로운 VR 헤드셋이 출시될 것으로 예상되며, 햅틱 및 디스플레이와 같은 주변 장치도 계속 발전하여 새로운 사용자 경험이 소비자 시장을 강화할 것으로 예상된다. 2023년 현재까지 플레이스테이션은 기존의 대규모 사용자 기반을 활용하여 VR2 헤드셋을 출시했으며, HTC는 하이엔드 소비자와 전문가 부문을 위한 바이브 XR 엘리트 헤드셋을 출시하여 긍정적인 평가를 받고 있다.
- 미래 기술의 도약을 예고** : 애플은 시선 추적, 손동작 인식, 끊김 없는 스크롤 등의 기술이 탑재된 비전 프로 AR/VR 헤드셋을 2024년에 출시할 예정이라고 발표했다. 애플은 인재와 자본을 바탕으로 5,000개 이상의 특허를 확보했고, 혼합 현실 영역에서 새로운 표준을 제시하고자 하고 있다.

불확실성

몰입형 현실 기술에 영향을 미치는 주요 불확실성은 다음과 같다.

- 소형화 및 경량화, 디바이스 내구성 강화, 센서 정밀도 향상, 사용자 편의성 증대, 발열 감소, 배터리 수명 연장 등을 위해서는 디바이스 하드웨어 및 소프트웨어 개선이 필요하다.

17) McKinsey, "Value creation in the metaverse", Oct 2022

디지털서비스 이슈리포트

2. 애플리케이션의 소비자 친화성과 확장성을 높이기 위해 경제적 접근이 필요하다.
3. 사용자 요구의 폭이 넓어지는 것은 아직 의문이다. 메타버스는 타겟으로 잡은 틈새 수요에서 더 광범위한 대중 시장 고객 사용으로 수요가 증가하는 지점에 도달하기까지 몇 년이 걸릴 것으로 보인다. 최종 소비자 가격대가 어떻게 진화되는지, 그것에 맞추어 특정 비즈니스 요구사항이 기능으로 채택되는지에 따라 영향을 받을 것이다.
4. 사용자 행동 추적과 관련된 보안 및 개인정보 보호 문제를 완화하는 것이 신뢰 구축에 매우 중요하다.
5. 고도로 통제된 환경 밖에서 시야를 제한하는 AR 및 VR 플랫폼의 사용을 고려할 때 안전 문제가 중요하다.

기술 트렌드를 이해해야 하는 이유

많은 기업이 디지털 트랜스포메이션을 진행하면서 클라우드나 인공지능, 머신러닝과 같은 기술을 도입하는 이유는 단순히 그것을 할 수 있는 역량이 있어서가 아니라 반드시 도입해야 하는 절대적 이유 때문이다. 새로운 디지털 기술은 기업의 민첩성, 혁신성, 확장성을 지원하기 전에 기업의 미래 생존을 결정하는 중요 요소이기 때문이다. 성공적인 기업이라면 클라우드에 있는 정보를 종합하고, 인공지능을 사용하여 데이터를 정보로, 지식으로 변환하고, 머신러닝을 사용하여 지식 간의 관계를 이해하고 비즈니스 문제에 학습을 적용하고, 의사 결정 및 자동화를 지원하는 차세대 소프트웨어 프레임워크를 통해 대규모 인사이트를 제공하는 "디지털 우선" 조직이 필요하다. 클라우드와 인공지능, 머신러닝은 비즈니스의 모든 기능 영역에 걸쳐 보편화되고 있다. 기술 트렌드를 이해하는 것은 디지털 우선 조직이 되기 위한 이니셔티브의 첫걸음이 된다. 맥킨지는 위에서 이야기한 5가지 외에도 양자 컴퓨팅, 디지털 신원, 재생에너지, 기후 기술을 2023년 주목해야 할 기술로 꼽았다.

업급된 모든 기술 하나하나가 우리 생애에서 가장 파괴적인 혁신을 가져올 것이다. 이러한 기술 트렌드를 이해해야 하는 이유는 세 가지이다.

1. 매출 성장

새로운 기술을 통해 잠재 고객을 더 효과적으로 타겟팅하고, 아웃리치 캠페인을 최적화하며, 리드의 우선순위를 정할 수 있다. 소셜 미디어 감정 마이닝, 광고 속성의 알고리즘 방식 선택, 마케팅 프로그램의 효과 측정, 고객 충성도 보장, 지능형 영업 추천을 통해 매출 성장을 기대한다.

2. 비용/운영 효율성

AI 기반 고객 센터 솔루션은 문제 해결 시간을 단축하고 고객 경험(CX)을 개선한다. 자연어

디지털서비스 이슈리포트

처리를 통해 고객은 번거로운 검색 과정 없이, 필요한 사항을 쉽게 말할 수 있다. 머신러닝 알고리즘은 질병의 진단과 치료를 가속화하고 정밀 의학을 지원한다. 산업용 엣지 컴퓨팅 모델은 기계의 고장 시기를 예측하고 예방적 유지보수를 권장하여 잠재적인 다운타임을 방지할 수 있다.

3. 리스크 완화

금융 기관은 대출 심사를 개선하고 리스크를 줄일 수 있다. 고급 사기 탐지 및 비정상적인 활동 발견을 통해 금융 범죄를 줄일 수 있다. 조직에 침해 예측 및 저지 능력을 향상하고, 제로 트러스트 보안 프레임워크로 급증하는 위협에서 보호해야 한다.

이와 같이 새로운 기술 동향의 올바른 이해를 통해 기업과 조직은 경제적 중요성과 성장 잠재력을 최대화하는 디지털 신기술 전략을 세울 수 있게 된다.

04 AI 기반 자율 운영 플랫폼 제공하는 비즈데이터

| 도안구 테크수다 편집장

공공자원 사업 매칭 서비스로 수요자와 공급자 동시 만족

비즈데이터는 ‘모든 비즈니스 의사 결정을 데이터로 돕자’를 모토로 데이터에 집중하는 회사다. 2014년에 회사를 만들고 7년째 사업을 전개하고 있다. 초기에는 재무와 경영 데이터를 주로 다뤘고 이를 바탕으로 데이터플랫폼 기술을 확보했다. 최근에는 제조와 환경 분야 데이터 영역까지 사업을 확장하고 있다. 최근에는 대한민국 4차 산업 리딩기업대상 ‘환경 AI 빅데이터’ 부문 대상을 받기도 했다.

데이터플랫폼과 인공지능, UI/UX를 통해 궁극적으로 사람의 개입을 최소화하는 자율운영서비스(AOS)가 가능한 솔루션을 제공하고 있다. 상·하수 관련 환경 분야의 그린인텔리전스(GI)와 기업 경영활동을 지원하는 비즈니스 인텔리전스(BI), 생산 관련 데이터를 실시간으로 수집, 분석해 데이터 기반 공장 경영 관리를 디지털화할 수 있도록 돕는 스마트 공장 구축, 데이터 관련 컨설팅 등 크게 4개의 영역에서 서비스를 제공한다.

디지털서비스 이용지원시스템에는 ‘지능형 능동 행정 공공자원 사업 매칭 서비스(PAMs)를 제공한다. 이 서비스는 공공지원사업 수집 데이터를 기반으로 공공기관 담당자가 지역 내 소상공인과 중소기업에게 최적의 신청 가능 공공정책자금 지원 사업을 추천 알고리즘을 통해 매칭 할 수 있도록 지원한다. 데이터 커버리지, 데이터 정제 기술, 매칭 노하우가 잘 결합되어 제공한다.

공공기관을 위해서 네이버 클라우드 플랫폼을 사용하고 있다.

다음은 비즈데이터와 나눈 일문일답.

1. 회사 소개를 부탁드립니다.

비즈데이터는 2014년 설립된 AI 빅데이터 전문기업입니다. 데이터와 AI를 활용해 자율운영서비스(AOS) 솔루션을 제공하고 있습니다. 제조와 환경 분야에 집중하면서 다양한 영역에서 생성되는 빅데이터를 가공·처리해 활용하도록 서비스를 제공하고 있습니다. 비즈니스 전략(BI) 분야를 주력으로 성장해 왔으며, 최근엔 수처리와 스마트 팩토리에 자율운영(GI) 플랫폼을 제공하고 있습니다. 디지털서비스 이용지원시스템에는 ‘지능형 능동행정 공공자원 사업 매칭 서비스(PAMs)를 제공합니다.

디지털서비스 이슈리포트

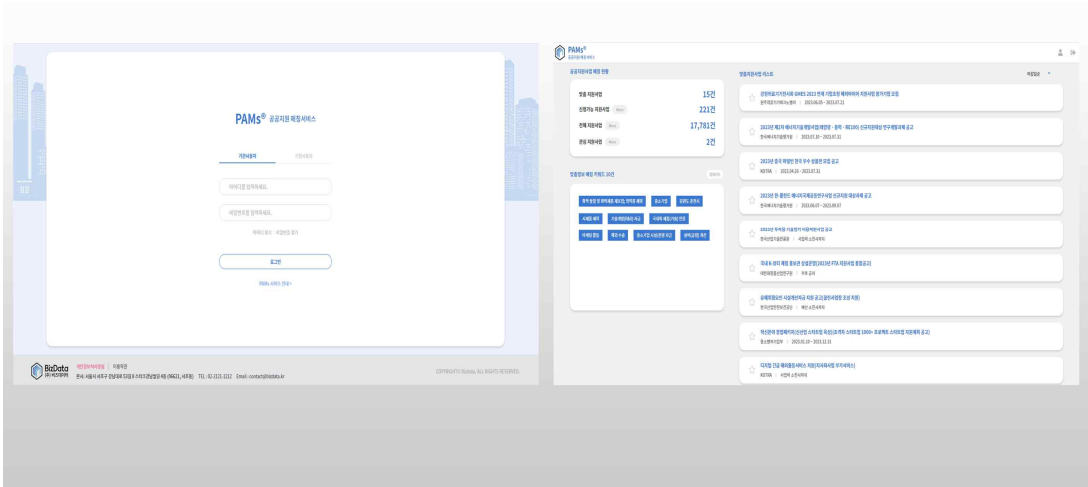


그림 9 비즈데이터의 공공지원 매칭 서비스 'PAMS'

팜스는 기초지자체 단위 지원 사업 데이터, 업종기초지자체 단위 지원 사업 데이터, 업종소분류 영역 데이터를 모두 제공해 서비스 인프라가 부족한 기초지자체, 지역 소재 기업지원기관, 이런 정보를 통해 공공 영역 프로젝트에 참여하려는 민간 기업이 활용할 수 있도록 돕고 있습니다.

2. 빅데이터 전문기업인데요, 제공하시는 솔루션과 제품은 무엇인가요?

비즈데이터는 AI 의사결정 플랫폼 기술을 중심으로 비즈니스 데이터를 활용하는 BI(Business Intelligence) 영역과 IoT(사물인터넷) 데이터를 GI(Green Intelligence) 영역에서 활용해 성장하고 있습니다.

또한 인공지능 수처리 자율운영서비스(AOS) 솔루션인 '스마트워터플랫폼'을 개발해 제공하고 있습니다. 스마트워터플랫폼은 하수처리 시설을 위한 자율운영 시스템으로 전해 응집, 세라믹 분리막 등 물리화학적 처리방식을 적용해 기존 하수처리 시설보다 데이터 수집이 용이하고 설치와 이동이 편리한 것이 특징입니다.

2022년부터 지속적으로 투자한 환경분야 AI 빅데이터 기술을 집약시켜 GI 사업부를 신설, 올해 GI 워터(water), GI 비전(Vision) 상표권을 등록해 사업을 전개하고 있습니다.

3. 최근 보내주신 보도 자료를 보니 수처리 분야에도 활동하고 계시던데요. 어떤 제품과 서비스를 제공하시나요?

디지털서비스 이슈리포트

환경 AI 솔루션 분야 전문 브랜드인 '그린인텔리전스'는 사물인터넷(IoT)과 AI 기술을 활용한 스마트시티 등의 인프라 자율운영 솔루션 브랜드입니다. 이러한 GI 자율운영 솔루션은 상하수도 수처리 설비 제어를 위한 기반 플랫폼이 됩니다. 발전소, 공장 등 스마트 팩토리의 실현을 위한 기반 플랫폼으로도 사용할 수 있습니다.

특히, 지능형 수처리 자율운영 솔루션 'GI Water'는 21년 정수장 수처리 약품 주입률 자동제어 솔루션으로 개발되었습니다만, AI 솔루션 적용을 점차 확대해 22년에는 EMS(Energy Management System)와 같이 정수장의 다른 분야에도 자율운영이 가능하게 되었습니다. 약품 주입률 자동제어 솔루션은 스마트 정수장 구축을 위한 핵심 공정으로 △최적 약품(응집제) 종류 및 약품별 주입률 산출 △약품 주입률 자동 제어 △분석 시스템 모니터링 △운영자 시각화 서비스 등으로 구성하여 개발한 솔루션입니다.

정수장 운영에서 가장 중요한 여름 장마철 시기에 무중단 자율운영을 지원, 정수장 운영자 검증을 통해 솔루션 성능의 신뢰성을 확보하는 성과를 거두어, 전국적으로 수처리 자율 운영체계가 확대되는데 밑거름 역할을 했다는 평가를 받고 있습니다.

4. '지능형 능동행정 공공지원 사업 매칭 서비스'에 대해서 설명 부탁드립니다.

많이들 알고 계시겠지만 “공공지원사업”이란 공공기관, 기업지원기관, 국책은행 등에서 우리 중소기업, 소상공인들을 지원하기 위한 사업을 말하며, 사업화 자금 이외에도 연구개발, 마케팅, 컨설팅, 인력 채용 등 다양한 분야를 지원하고 있습니다.

하지만, 중소기업(소상공인) 처지에서는 '풍요속의 빈곤'이라고, 지원 사업 내용이 복잡해 정보탐색과 사업 신청 과정에 많은 애로사항을 호소하고 있습니다. 공공기관 담당자 입장에서는 관내 기업의 애로사항을 지속적으로 파악하고, 해당 기관(지자체) 공공지원사업뿐만 아니라 타 기관, 중앙부처 지원 사업과 연계해서 지원하고자 하는 요구가 있습니다.

비즈데이터의 '지능형 능동행정 공공지원 사업 매칭 서비스(PAMs : 팜스)'는 공공지원사업 수집 데이터를 기반으로 공공기관(지자체, 기업지원기관) 담당자가 본인이 담당하는 공공지원사업 대상이 될 수 있는 관내 중소기업(소상공인)을 신용정보회사 기업 데이터에서 추출, 매칭 할 수 있으며, 타 공공기관의 지원 사업을 벤치마킹할 수 있습니다. 또한, 중소기업(소상공인) 정보를 기반으로 해당 기업의 최적 공공지원 사업을 추천할 수 있도록 지원하는 서비스입니다.

디지털서비스 이슈리포트

5. 이 서비스를 통해 기업 혹은 행정 기관들이 얻은 이점은 무엇인가요

팜스(PAMs) 활용을 통해 공공기관 담당자들이 중소기업(소상공인) 지원 사업을 기획할 때, 타 기관(지역)의 유사 공공지원사업을 벤치마킹할 수 있고, 신규 공공지원사업의 대상이 될 수 있는 관내 기업을 발굴할 수 있습니다. 이를 통해 잠재 수요를 발굴하는 행정서비스가 가능합니다. 또한, 공공지원사업의 대상기업을 지속적으로 발굴하고 지자체-중앙정부 지원 사업 통합정보를 선제적으로 제공함으로써 능동적 행정서비스를 구현할 수 있을 것으로 판단합니다.

예를 들어, 시청 기업지원과에 '지자체 청년 창업지원사업' 신청 서류를 제출하기 위해 방문한 청년 창업자에게 담당 공무원이 해당 창업기업의 특성 정보를 팜스(PAMs) 서비스에 입력해 중앙부처/광역지자체의 신청 가능한 타 지원 사업 정보를 제공함으로써, 다양한 사업화 자금을 확보할 수 있도록 지원할 수 있습니다.

6. 구체적인 사례 소개 부탁드립니다.

현재, KB국민은행 브릿지서비스, 대구은행, 전북은행 등과 네이버파이낸셜에 서비스 개발과 정책자금 데이터 서비스를 제공하고 있습니다. 이번, 공공기관용 공공지원 사업 매칭 서비스를 SaaS 형태로 제공하게 되었고, 고양시청 등 기초지자체를 중심으로 서비스를 활용할 예정입니다.

7. 타사 서비스에 비해서 갖는 경쟁력은 무엇인가요

첫째로는 수집 데이터 커버리지입니다. 팜스(PAMs)의 공공지원사업 데이터는 통합공고사이트, 기초/광역지자체 등 국내 공고 정책자금의 90% 커버리지 수준을 수집하고 있으며, 매칭 정확도 향상을 위해 공공지원사업의 업종(KSIC코드) 소분류를 적용하고, 지원 사업 형태를 25개 분류체계로 분석하고 있습니다.

둘째로는 데이터 정제 기술입니다. 공공지원사업 공고문은 비정형/반정형 텍스트로 비즈데이터는 머신러닝이 적용된 정제/추출(내용정제, 태깅정제) 모듈을 이용해 정형화 데이터로 정제하고, 정합성 보증을 위해 데이터 검수자의 최종 확인을 거친 후 데이터베이스(DB)화하고 있습니다.

셋째로는 공공지원사업 매칭 노하우입니다. 자사는 공공기관 경력 임직원들 중심으로 팜스(PAMs) 서비스를 개발해 왔습니다. 또 다수의 금융권, 공공기관 공공지원사업 매칭 프로젝트 참여를 통해 당사의 노하우를 기반으로 매칭 알고리즘을 지속적으로 고도화하고 있습니다.

넷째로는 공공기관의 서비스 도입을 위한 필수 인증을 획득했습니다. 팜스(PAMs)는 공공기관에서 클라우드

디지털서비스 이슈리포트

서비스 도입 시 가장 우려되는 요소인 보안 적합성을 판단하는 ‘클라우드 보안인증(CSAP)’을 획득하고, 클라우드 서비스의 신뢰성과 안정성을 보장하기 위한 ‘클라우드 서비스 확인제’와 ‘클라우드 플랫폼 PaaS-TA 호환성 확인제’도 획득했습니다. 이를 바탕으로, 한국지능정보사회진흥원(NIA)의 ‘디지털서비스 이용지원시스템’에 등록했습니다.

8. 제품 개발과 서비스 제공이 더 확산되기 위해 정부가 더 노력해야 하는 영역은 어떤 분야인가요.

법적, 제도적 기반이 마련되어 있지만, 공공기관에서는 여전히 정보보안 등의 이유로 클라우드 서비스(SaaS)보다는 온-프레미스형 구축을 선호하는 경향이 있습니다. 당사의 팜스(PAMs) 서비스와 같은 클라우드 서비스의 공공분야 확산을 위하여 SaaS 형 서비스를 행정예 선도적으로 도입하는 공공기관(지자체)에 대한 인센티브가 확대되어야 할 필요가 있습니다.

8. 올해 관련 사업 분야의 목표도 궁금합니다.

비즈데이터는 환경과 에너지·안전 등의 AI 자율운영 서비스를 위해 다양한 분야에서 전문기업들과 파트너십을 구축, 스마트 워터 플랫폼을 공동으로 개발·서비스 중에 있습니다. 새로운 수처리 시스템 도입으로 AI 자율운영, 탄소 중립, 에너지 최적화를 통해 안정적이고 깨끗한 물관리 서비스를 제공할 수 있도록 노력할 예정입니다.

